Crowd Motion Capture

Nicolas Courty, Thomas Corpetti

Crowd Motion Capture

Nicolas Courty
Samsara/VALORIA/Université de Bretagne Sud
Campus de Tohannic
56000 Vannes, France
Nicolas.Courty@univ-ubs.fr

Thomas Corpetti
COSTEL/Université de Haute-Bretagne
Place du Recteur Henri Le Moal
35043 Rennes, France
Thomas.Corpetti@uhb.fr

ABSTRACT

In this paper a new and original technique to animate a crowd of human beings is presented. Following the success of data-driven animation models (such as motion capture) in the context of articulated figures control, we propose to derivate a similar type of approach for crowd motions. In our framework, the motion of the crowds are represented as a time series of velocity fields estimated from a video of a real crowd. This time series is used as an input of a simple animation model that "advect" people along this time-varying flow. We demonstrate the power of our technique on both synthetic and real examples of crowd videos. We also introduce the notions of crowd motion editing and present possible extensions to our work.

Categories and Subject Descriptors

I.3.7 [**Computer Graphics**]: Three-Dimensional Graphics and Realism - Animation; I.6.8 [**Simulation and Modeling**]: Types of Simulation - Animation

Keywords

Crowd simulation, data-driven animation, vision-based techniques for computer animation, motion estimation

1. INTRODUCTION

Crowds of people exhibit particular and subtle behaviors whose complexity reflects the complex nature of human beings. Animating crowds are usually a tedious task for animators who rely most of the time on simulation models. Achieving particular effects with global emergent crowd behaviors from such models appear to be very tricky, since animators usually only have control on specific characteristics of the crowd members. In the context of human-like figures animation, huge progress have been observed with the use of motion capture. Using motions acquired from real performers is now common, and modifying them through a variety of editing operations brings substantial benefits in terms of control and realism in the produced animation. The aim of our technique is to adapt such a methodology in the context of crowd animation. This raises an opened set of questions:

- what kind of descriptors will best represent crowd motions ?
- how to capture/estimate them from real situations ?
- how to apply/use them as an input for an interactive, parameterizable animation system ?
- and finally is it possible to edit those captured motions in order to produce new motions ?

We chose to use crowd videos as input data because they are relatively easy to capture and do not require invasive techniques such as markers that are impractical in the presence of thousands of individuals. At first it sounds temptating to track singular pedestrians into the flow of people. We argue that, despite the technical difficulties to achieve such a task, this approach fails in describing the crowd behavior as a whole (but may as well succeed in characterizing individual specific behaviors). Conversely, our framework is based on the assumption that the motions of individuals within the crowd is the expression of a continuous flow that drives the crowd motion. *This assumes that the crowd is dense enough so that pedestrians are considered as markers of an underlying flow.* In this sense, our method is more related to macroscopic simulation models (that try to define an overall structure to the crowd's motions) rather than microscopic models (that define the crowd's motions as an emergent behavior of the sum of individual displacement strategies).

Overview of the data-driven animation system. Our methodology relies on an analysis/synthesis scheme which is depicted in Figure 1. First, images are extracted from a video of a real crowd. From all the pairs of successive images a vector field is computed through a motion estimation process. The concatenation of all these vector fields represent a time series which accounts for the displacement of the whole crowd along time. This ends up the analysis part. The synthesis of a new crowd animation is done by advecting particles (the pedestrians) along this time varying flow.

This paper is divided as follow: Section 2 presents related work in the context of crowd simulation as well as motion estimation. Section 3 deals with the estimator used in our

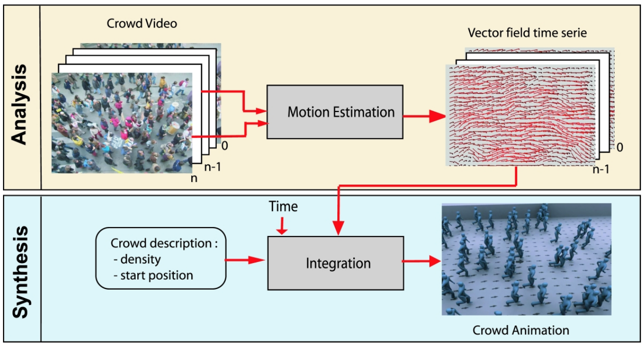

Figure 1: Overview of the whole process

methodology, and Section 4 presents the integration of the motion descriptor in a crowd animation controller, as well as possible ways to edit the captured motions. The last two sections present results obtained with our method along with a conclusion and perspectives for our work.

2. STATE OF THE ART

Data-driven animation techniques usually require means to extract information from the real world. In the presence of rigid motions such as character motions, marker based techniques bring the best results because only a few motion parameters have to be considered. In the context of non-rigid motions such as facial parameters, fluids flows, animal gaits/behaviors or tree structures, video-based techniques have already been considered in the literature [10, 4, 15, 16, 14]. Simulating crowds from real videos is closely related to this topic, but yet is different because of the choice of the motion descriptor (time varying vector fields). Such a description has been previously considered in a general animation system, *Flow tiles* [11], but the author did not consider that the flow tiles could be designed from real data. In [9], Brogan *et al* succesfully performed statistics on captured human walking paths and derived from them a new path planning model. In order to motivate a data-driven animation scheme for crowd animation we first present the state of the art works in crowd animation, then we introduce some general issues in the motion estimation problem.

2.1 Crowd simulation

Crowd behavior and motion of virtual people have been studied and modeled for several purposes ranging from movie production to the simulation of emergency situations or the design of buildings and open-spaces. The state of the art in human crowd behavioral modeling is large and can be classified in two main approaches: microscopic and macroscopic models. The models belonging to the first category describe the time-space behavior of individual pedestrians whereas the second category describe the emergent properties of the crowd. The simplest models of microscopic simulation are based on cellular automata [7, 6]. Dynamic systems have also been considered by Helbing in his popular social force model [17]. It consists in expressing the motion of each pedestrian as a result of a combination of *social forces*, that repel/attract pedestrians toward each others. It has been shown that this model generates realistic phenomena such as arc formations in exits or increasing evacuation time with increased desired evacuation velocities. It has been extended to account for individualities [8] or the presence of toxic gases in the environment [12]. More complex models consider each member of the crowd as autonomous pedestrians endowed with perceptive and cognitive abilities [25, 30, 26, 28]. Those models exhibit a variety of results depending on the quality of the behavior design. Such models may lead to incorrect emergent global behaviors. Those difficulties can be avoided by using a continuum formulation [20, 31]. Equations using the concepts of fluid mechanics have been derived in order to model such approach of human crowds. Those approaches rely on the assumption that the characteristic distance scale between individuals is much less than the characteristic distance scale of the region in which the individuals move [20]. As such, the density of the crowd has to be taken into account for those models to be pertinent. Finally several hypotheses on the behavior of each members of the crowd lead to partial derivative equations governing the flow of people.

Although crowds are made up of independent individuals with their own objectives and behavior patterns, the behavior of crowds is widely understood to have collective characteristics which can be described in general terms. Though, macroscopic models may lack of subtleties and often rely on strong hypotheses (notably on density). Our framework propose to capture this global dynamic from real crowd video sequences. This imposes the use of motion estimation techniques.

2.2 Motion estimation

Tracking a very short number of people and using their motions in a computer animation system has previously been investigated by Somasundaram and Parent in [29]. When a crowd is dense enough, usual tracking systems like Kalman filters or stochastic filtering [13, 33, 32] generate large state space that yields a computationally expensive problem. It is then necessary to use alternative methods to obtain the information on the dynamics of the crowd in order to characterize its behavior. The idea of using optical flow to estimate crowd motions has recently drawn attention in the context of human activity recognition [1]. This section investigates the most popular ways to estimate the motion from image sequences.

From the panel of available approaches, almost all of them are based on the well-known Optical-Flow Constraint Equation (ofce): it assumes the visible points conserve roughly their intensity in the course of a displacement. This reads:

$$\frac{dE(\mathbf{x},t)}{dt} = \frac{\partial E(\mathbf{x},t)}{\partial t} + \boldsymbol{\nabla} E(\mathbf{x},t)\cdot \mathbf{v}(\mathbf{x},t) \approx 0, \qquad (1)$$

where $\mathbf{v}(\mathbf{x},t) = (u,v)^T$ is the unknown velocity field at time t and location $\mathbf{x} = (x,y)$ in the image plane Ω, $E(\mathbf{x},t)$ being the image brightness, viewed for a while as a continuous function. It is known that this assumption is subject to the aperture problem: in homogeneous areas, the latter equation is undetermined and there exist an infinity of solutions. A number of approaches have hence been proposed to cope this problem and to obtain the optical flow under the ofce constraint (see for instance [24] for a survey). In the following, we list the most popular ones.

Correlation: these techniques are based on the research of a window that is similar (w.r.t the ofce) to a window centered at position $\mathbf{x}$. The displacement associated corresponds to a peak of the correlation surface. Very efficient implementations of such approaches can be obtained in the Fourier space but these techniques suffer from a lack of consistency (from a spatial point of view) and some known limitations of such approaches [2] prevent from a comfortable use.

Parametric models: the idea is to cope the aperture problem by using a parametric model Θ to represent the velocity in the image ($\mathbf{v}(\mathbf{x}) = \Theta\mathbf{x}$). The model can be obtained by minimizing:

$$\Theta = \min_{\Theta} \int_{\Omega} \left(\frac{\partial E(\mathbf{x}, t)}{\partial t} + \boldsymbol{\nabla} E(\mathbf{x}, t) \cdot \Theta\mathbf{x} \right)^2. \qquad (2)$$

Such techniques are very efficient when one has a reliable a priori knowledge on the flow to estimate but this representation strongly limits the possibility of estimating the motion in a number of various situations.

Lucas & Kanade: Lucas and Kanade [22] assumed that the unknown optic flow vector is constant within some neighborhood of size n. Hence, the motion field $\mathbf{v}$ can be estimated by minimizing the function:

$$\int_{\Omega} K_n * \left(\frac{\partial E(\mathbf{x}, t)}{\partial t} + \boldsymbol{\nabla} E(\mathbf{x}, t) \cdot \mathbf{v}(\mathbf{x}, t) \right)^2 \qquad (3)$$

where the Gaussian Kernel K_n tends in fact to eliminate homogeneous areas. The main drawback of this approach is its incapacity of estimating local measurements.

Optical Flow: In the seminal work of Horn & Schunck [18], the authors proposed to solve the ofce by adding a spatial regularization constraint. The motion field is then obtained by minimizing:

$$\begin{aligned} &\int_{\Omega} \left(\frac{\partial E(\mathbf{x}, t)}{\partial t} + \boldsymbol{\nabla} E(\mathbf{x}, t) \cdot \mathbf{v}(\mathbf{x}, t) \right)^2 \\ &+ \alpha \int_{\Omega} f^2(\mathbf{v}), \end{aligned} \qquad (4)$$

where f is a smoothing function and α a parameter to fix. In the original work, the authors proposed a first order spatial smoothing term: $f(\mathbf{v}) = \|\boldsymbol{\nabla} u\|^2 + \|\boldsymbol{\nabla} v\|^2$ that assumes a spatial coherency of the motion field. This penalty term is based on the idea that all points of the same rigid object have a similar motion. Unfortunately, in presence of discontinuities, this term is not adapted as it smooth out the frontiers. The two last decades, many authors have proposed some regularizations of the type $f(\mathbf{v}) = g(\|u\|) + g(\|v\|)$ where g is a robust function to define. Its objective is to extract a piecewise smooth motion field while preserving the discontinuities (see for instance [5, 23]).

3. CROWD ESTIMATION AND REPRESENTATION

In this section we present our motion estimation technique along with necessary post-processing treatments that enhance both the quality of the signal as well as its compactness.

3.1 Motion estimation

It has been shown in [2] that a penalty term based on a first order smoothing $f(\mathbf{v}) = \|\boldsymbol{\nabla} u\|^2 + \|\boldsymbol{\nabla} v\|^2$ is equivalent, from a minimization point of view, to the minimization of:

$$\int_{\Omega} \mathrm{div}^2(\mathbf{v}) + \mathrm{curl}^2(\mathbf{v}) d\mathbf{x} \qquad (5)$$

where $\mathrm{div}(\mathbf{v}) = \partial u/\partial x + \partial v/\partial y$ and $\mathrm{curl}(\mathbf{v}) = \partial v/\partial x - \partial u/\partial y$ are respectively the divergence and the vorticity of the flow. Therefore, a classic smoothing term minimize areas that exhibit high values of divergence and/or vorticity. Such kind of motions descriptors are issued from fluid mechanics and are also very useful to characterize a crowd phenomenon: the divergence represents a dispersion while the vorticity is linked to a walk around an obstacle. Some studies have actually proved that a crowd dense enough has sometimes a behavior that can be explained by some fluid mechanics laws [20]. It is then of primary interest to integrate such prior knowledge in the optical flow to obtain a technique devoted to crowd motion. In particular, it is important to measure precisely the divergence and the vorticity from real data to perform an accurate synthesis of the observed phenomenon. In this paper, we then prefer to rely on a *div-curl* smoothing term that extract more precisely those two components of the flow [2]. The new smoothing function f is then $f(\mathbf{v}) = \|\boldsymbol{\nabla}\mathrm{div}\ \mathbf{v}\|^2 + \|\boldsymbol{\nabla}\mathrm{curl}\ \mathbf{v}\|^2$. Such a penalization tries to preserve areas that contain high values of divergence and/or vorticity.

3.2 Motion post-processing

In practice, the real sequences of crowds are composed of a high number of images (between $250(\approx 10sec)$ and $1500(\approx 1min)$). This results in an important number of instantaneous dense motion fields. Under the assumption that the pedestrians are some "tracers" of the underlying flow that transport them, the information contained by all the instantaneous motion fields is redundant. Moreover, some locations of the images are pedestrian-free (only the background is visualized); in the time series this results in some noisy areas where the corresponding displacements are computed only with the help of the spatial-regularization. In such situations, from an image to an other, there is no warranty that the flow is consistent from a temporal point of view. This results in noisy time series, as depicted in the green signal of figure 5. According to these remarks, it is then of primary interest to *i)* reduce this huge amount of data and *ii)* de-noise the time series. To that end, we suggest to exploit some properties of the Fourier transformation.

Fourier transformation Following the Fourier transform, any discrete signal of size T can be represented into a series of T growing frequency components. The lowest frequency represents the mean of the signal; the lower frequencies represent its small variations; higher frequencies are linked to high variations whereas very high values represent noises. It is then important to reconstruct a signal by removing its higher frequencies in order to de-noise it. Concerning the reduction of dimension, the figure 2 represents the error of reconstruction of a complete sequence of motion fields in function of the number of frequencies used to reconstruct it. It is shown than if we take a few number of frequencies (less than 10%), the reconstruction is fairly accurate.

4. DATA-DRIVEN CROWD ANIMATION

Once the time series of motion fields has been computed, it is possible to consider this information as input data for

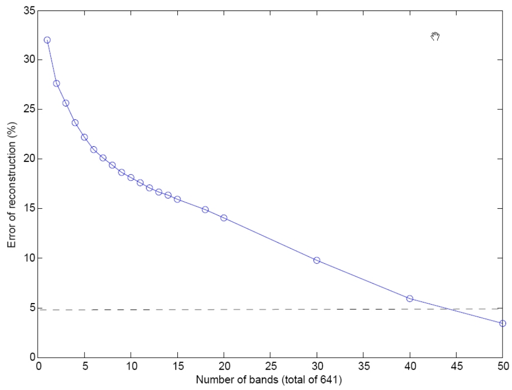

Figure 2: Reconstruction error. In this example, the original signal contain 641 frequencies. The reconstruction error continuously decreases and if one takes 50 frequencies (less than 8%), the error becomes lower than 5%

a data-driven crowd animation system. We first present in this section a simple model to compute crowd motions from these time series (*Motion synthesis*). We then exhibit two possible techniques to edit/modify those data (*Motion editing*).

4.1 Motion synthesis

Let us first recall that the computed velocities correspond to a velocity *in the image space*, and our goal is to animate pedestrians *in the virtual world space*. Given the position of such a person in the virtual world, it is possible to get the corresponding position in the image frame along with a camera projection model. Parameters for this projection can be obtained exactly through camera calibration. We have considered as an approximation of this model a simple orthographic projection in the experiments presented in the result sections. This assumption holds whenever the camera is sufficiently far away from the scene. Once this projection has been defined, animating individualities which constitute the crowd amounts to solve the classical following differential equation (with $x(t)$ the position of a person in the image frame at time t) :

$$\frac{\partial \mathbf{x}}{\partial t} = \mathbf{v}(\mathbf{x}(t), t) \tag{6}$$

equipped with appropriate initial condition $x(0) = x_0$ which stands for the initial positions of the individual in the flow field. In our framework we have used the classical 4-th order Runge Kutta integration scheme, which allows to compute a new position $x(t+1)$ given a fixed time step with an acceptable accuracy. This new position is then projected back in the virtual world frame. This process is depicted in Figure 3.

4.2 Motion editing

Continuous crowd motion. Assuming that only a short video sequence of a crowd is available, we present here a technique to generate a continuous and aperiodic animation of crowd. Following the work on video textures [27] (which was later used in the motion graph techniques for body motions [21]), the idea is to generate transitions be-

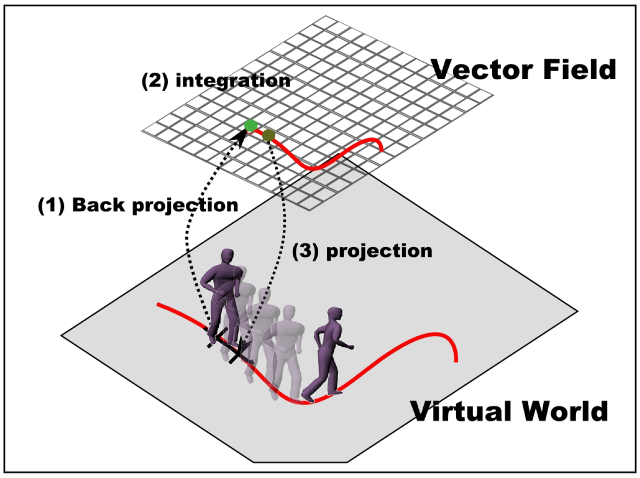

Figure 3: Motion synthesis from flow field. The position of the crowd's member is projected onto the flow (step 1), the integration is performed in the image frame (step 2) and then the new position is projected back in the virtual world frame (step 3).

tween different instants in the vector field time series based on a similarity criterion. In order to compare different instantaneous velocity fields we propose to use, as suggested in [3], the angular error:

$$d(\mathbf{U}, \mathbf{V}) = \int_{\Omega} \arccos\left(\frac{< \mathbf{u}(\mathbf{x}), \mathbf{v}(\mathbf{x}) >}{||\mathbf{u}(\mathbf{x})||||\mathbf{v}(\mathbf{x})||}\right) d\mathbf{x} \tag{7}$$

where $\mathbf{U}$ and $\mathbf{V}$ are two vector fields of same dimensions, $\mathbf{u}(\mathbf{x}) = (\mathbf{U}(\mathbf{x}); 1)$ and $\mathbf{v}(\mathbf{x}) = (\mathbf{V}(\mathbf{x}); 1)$ correspond to the homogeneous coordinates of the fields vector elements (vectors of dimension 3). The rest of the procedure is then identical to [27]: the distance matrix is filtered with an n-tap filter that magnifies similarities upon a small time window. From this matrix are derived probabilities of transition from one instantaneous field to another. Figure 4 shows an example of such matrices built upon a 3 seconds small crowd video.

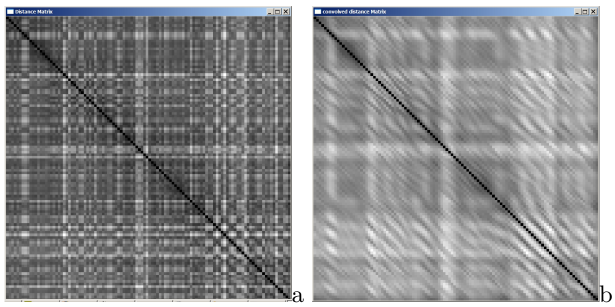

Figure 4: distance matrix; (a): distance matrix using an angular distance between each vector field in the time series; (b) filtered matrix with an 8-tap filter

Perspectives. There exists a variety of potential editing operations for crowd motions. As perspectives for this work we sketch two of them:

- **Changing the crowd style**. Using the Fourier Transform may also help in separating different patterns in the crowd behavior, depending on the classes of frequency used for the reconstruction. A signal composed with only few low coefficients actually corresponds to a stable and regular behavior. Conversely, manually adjusting the scale of high frequencies for instance is likely to produce a more "nervous" pattern in the crowd motions, thus giving the animator a potential control over the crowd "style". Examples of reconstruction using either low or high frequencies are given in figure 5

- **Tiling crowd patterns**. Given several available motion time series one could possibly envisage to combine them spatially in the same virtual environment (one entrance and one obstacle acquired separately for example). The problem is here to define a proper juxtaposing or fusion tool that will preserve the interesting features of the flow. Let us finally note that this spatial registration can also be done in the time domain (how to blend two series together ?). Dynamic Time Warping techniques stand as good candidates for those purposes.

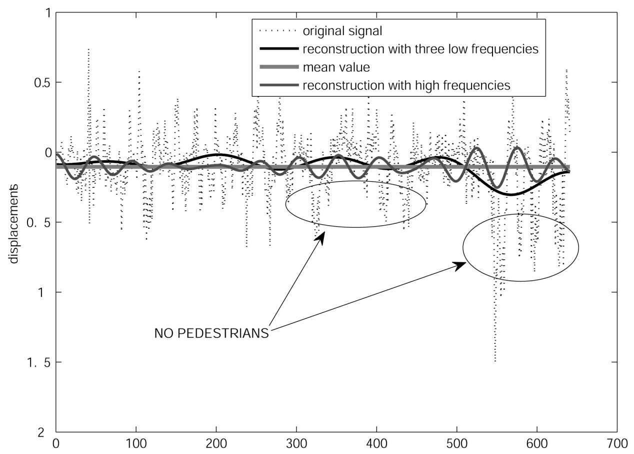

Figure 5: Illustration of the Fourier Transform. dashed lines: an initial time series; line in gray: its mean value (reconstruction with one frequency); black signal: reconstruction with 3 frequencies; the resulting filtered time series is smooth and corresponds to a quiet displacement; gray signal: the de-noised signal (very high frequencies removed) reconstruct with only high frequencies; this is linked to a stressed behavior. The black ellipses are some examples of pedestrian-free areas.

5. RESULTS

Our approach has been first tested on a synthetic crowd sequence to validate the theoretical part of our work. We have also used real sequences to handle real cases. Those results are presented in this section. Concerning the post-processing, each time series of motion fields has been represented with only 10% of its lowest frequencies. This choice is based on the graph of figure 2.

5.1 Synthetic example

The synthetic sequence represents a continuous flow of human beings with an obstacle (a cylinder named $\mathcal{C}$) in the middle of the image. It has been generated using the classical Helbing simulation model [17].

Motion estimation process In this situation, the area $\mathcal{C}$ is subject to the aperture problem: any vector inside $\mathcal{C}$ is a reliable candidate. The motion estimation technique being equipped with a spatial regularization, we need to apply a specific process to prevent from incoherent motions that may occur inside the cylinder. To that end, we completely blurred this area from an image to an other, so that the OFCE constraint is verified nowhere inside $\mathcal{C}$. We applied a robust M-estimator to the OFCE (first term of relation (4)) instead of a quadratic penalization. Such functions, arising from robust statistics [19], limit the impact of the locations where the brightness constancy assumption does not hold. As a consequence, this area is not taken into account by the observation term of the estimation process. The motion fields estimated outside the cylinder are then not disturbed by the ones inside $\mathcal{C}$. This is illustrated in Figure 6. We present 4 images of the sequence in Figure 6(a–d); an estimated motion field is depicted in Figure 6(e) and a zoom of the cylinder area with and without the specific treatment is presented in Figure 6(f) and 6(g) respectively. Concerning the smoothing parameter α, its value was set to 250 (this corresponds to the dynamics of the brightness function).

Animation Some images of the crowd animation synthesis are shown on Figure 6(i–l). The animation was generated thanks to a Maya plugin which defines a crowd as a set of particles and performs the synthesis described in section 4. As expected, the virtual crowd moves in accordance with the underlying motion and the obstacle is correctly managed. This first example proves the ability of the proposed approach to synthesize a coherent motion from an estimated motion field. Let us now apply this technique to real data.

5.2 Real data

We present the results obtained on two real sequences. Both data have been acquired with a simple video camera with an MPEG encoder. The motion estimation approach was applied without any particular process, with a smoothing parameter $\alpha = 250$.

Strike sequence The first real sequence is a video representing a strike. All pedestrians are walking in the same direction. Two images of the sequence can be seen on Figure 7 (a-b). In Figure 7 (c-d), we present the synthetic crowd animation obtained superimposed on the estimated motion field. One can observe that the resulting crowd animation respect the initial yet simple pedestrian behaviors. The real scene can then be reproduced with accuracy while maintaining, despite the regularity of the flow, a particular diversity in the pedestrian trajectories.

Entrance sequence The second real sequence shows a crowd entering a stadium. This example is very worthwhile since a variety of phenomena are present: a continuous flow at the beginning followed by a compression of some peoples in the left part of the images. In addition, the limit of the door is an obstacle that creates two opposite fluxes and that generates a vortex in the motion fields. Four images of the sequence are displayed in Figures 7 (e–h).

The animation is represented in Figures 7 (i–l). Figures 7 (m–p) are focused on the region that exhibits opposite mo-

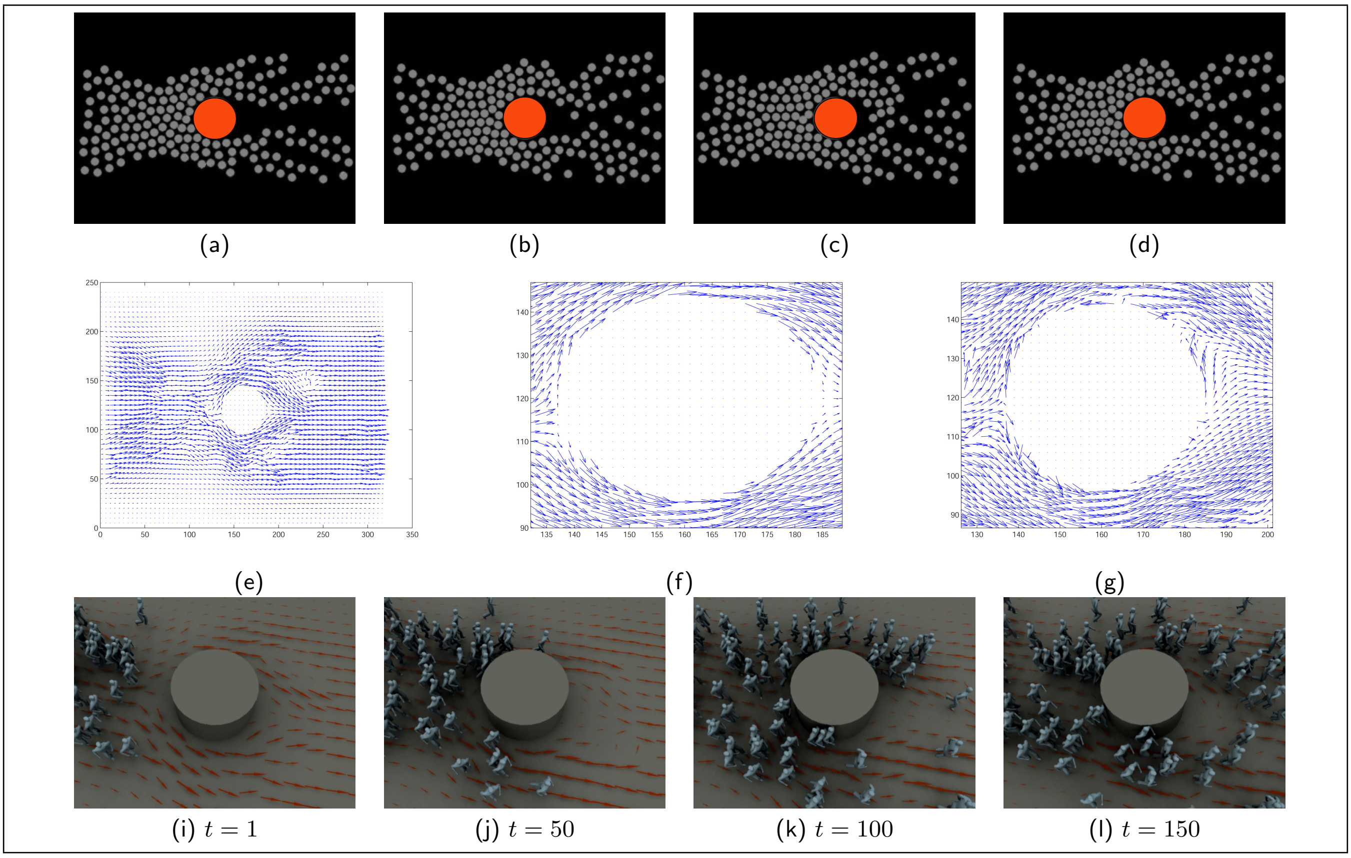

Figure 6: Experimental results on synthetic data: (a,b,c,d): Sequence #1: simulation of a crowd flow with a cylindric obstacle; (e) the estimated motion field;(f) motion near the cylinder estimated with a special care of this no-data area and (g) same as (f) but without a specific treatment for the cylinder. One can see that the motion near the cylinder in (g) is not totally coherent. (i,j,k,l): images of the simulated crowd with the synthetic flow.

tions. One can see that this complex behavior has been correctly captured and re-synthesized. This is very stimulative regarding the possibilities of this approach to manage complex flows. The next step of the process will be to extract the different behaviors (compression, rotation for instance) and to synthesize them independently.

Quantitative values The table 1 shows some details on the size, the time computation (generated on a $2GHz$ PC with $1Go$ RAM) and the quality of the compression for the three sequences. We also added the space required by the whole motion field and the compressed one. Concerning the animations, they were generated at interactive framerates (60 frames per second). It can be observed that our approach is thus suitable for real-time applications and does not require a lot of memory.

	seq #1	seq #2	seq#3
1	$(80, 60)$x214	$(80, 60)$x274	$(80, 80)$x641
2	3 min 35 s	4 min 35 s	$\approx$ 11 min
3	19.2 Mo	24.7 Mo	76.4 Mo
4	1.92 Mo	2.47 Mo	7.64 Mo
5	93.4%	97.3%	97.5%

Table 1: Numerical aspects. 1^{st} line: size of the sequences (lin,col)$\times$ t; 2^{nd} line: CPU time for the motion estimation; 3^{rd} line: uncompressed size of the motion field; 4^{rd} line: compressed size; 5^{th} line: percentage of the information in the compressed field.

5.3 Discussion

Our technique has been applied with success to reproduce the observed scenes. Nevertheless, we have isolated two limitations to our approach:

- the quality of the generated animation is linked with the initial density of the crowd members. In this sense, it is the role of the animator to design an initial crowded situation that is similar to the video conditions, though a certain latitude is possible. This point requires to our opinion more investigation,
- this analysis/synthesis scheme may not keep people away from colliding between each other. This point could be solved by mixing the *a priori* information acquired by motion estimation as an input parameter of a classical dynamic system such as Helbing's model [17] or the more recent Continuum crowd's model [31]. Those aspects are part of our future works.

6. CONCLUSION

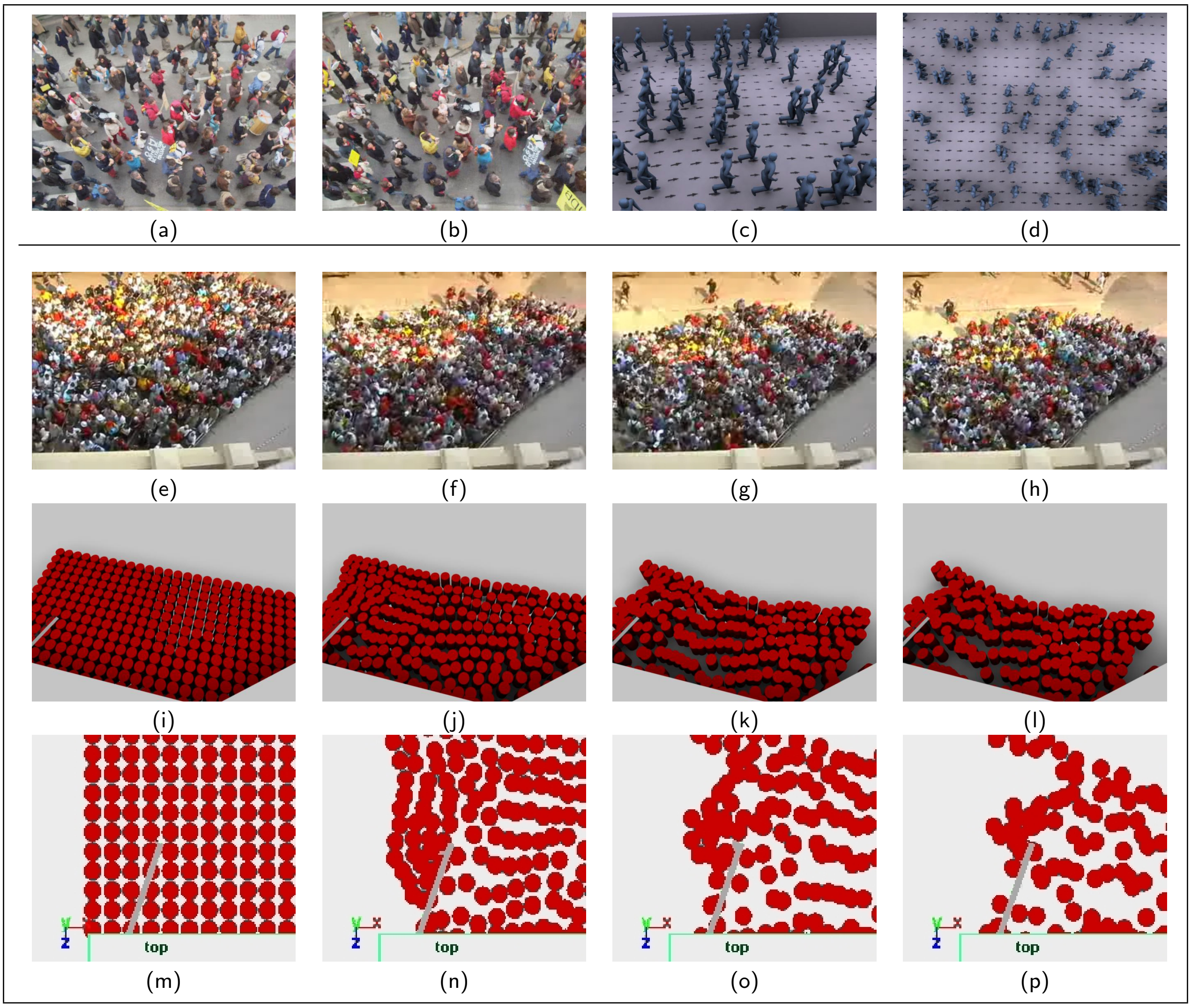

Figure 7: Experimental results on real data: (a,b): Sequence #2: strike video taken from above; (c,d) images of the resulting animation; (e,f,g,h) Sequence #3: video of a crowded entrance; (i,j,k,l) images of the resulting animation; (m,n,o,p) close-up on a remarkable zone where two opposite fluxes of people are juxtaposed

In this paper an extension of the concept of motion capture was proposed to the domain of crowd animation, based on the analysis of video sequences of crowd. Our framework relies on *i)* a specific motion information process applied on the video images and *ii)* a data-driven animation scheme that allows to generate animation of crowds from this input information. We applied the presented method on both synthetic and real examples. In our opinion the results are convincing and we believe that this method could be efficiently used by animators to produce realistic crowd effects from sample situations. We also think that this approach can be used to compare the results of simulation models to real situations according to their underlying flows, thus providing a quantitative evaluation framework (rather than the traditional qualitative one). Motivated by our experiments, three points are envisaged to continue this work:

- concerning the analysis part, the motion estimation process can be improved by introducing more specific spatio-temporal models of continuous crowd formulation like [20]. This can be done through the use of the optical flow framework but it can also be relevant to explore the data-assimilation possibilities used for instance in meteorology,
- concerning the synthesis part, it is possible to consider more sophisticated animation system by introducing the *a priori* information obtained from the analysis part in dynamic systems such as [17, 31],
- finally we plan to continue the analogy with motion capture and propose new editing methods that will make it possible to slightly change/control the intial captured motions.

7. ACKNOWLEDGMENTS

The authors wish to thank Amel Achour and Gwendal Jabot for their contributions to the project, and also reviewers for pointing out missing references in our work.

8. REFERENCES

[1] E. Andrade, S. Blunsden, and R. Fisher. Modelling crowd scenes for event detection. In *Int. Conf. on Patt. Recognitiion, ICPR 2006*, pages 175–178, 2006.

[2] Anonymous. Not cited due to anonymous review process. *IEEE Trans. Patt. Anal. Machine Intell.*

[3] J. Barron, D. Fleet, and S. Beauchemin. Performance of optical flow techniques. *Int. J. Comp. Vis.*, 12(1):43–77, 1994.

[4] K. Bhat, S. Seitz, J. Hodgins, and P. Khosla. Flow-based video synthesis and editing. *ACM Tran. on Graph., special issue, Proc. ACM SIGGRAPH 2004*, 23(3):360–363, 2004.

[5] M. Black. Recursive non-linear estimation of discontinuous flow fields. In *Proc. Europ. Conf. Comp. Vis.*, pages 138–145, Stockholm, Sweden, 1994.

[6] V. Blue. Cellular automata microsimulation for modeling bi-directional pedestrian walkways. *Transp. Res., Part B: Meth.*, 35(3):293–312, Mar. 2001.

[7] E. Bouvier, E. Cohen, and L. Najman. From crowd simulation to airbag deployment: particle systems, a new paradigm of simulation. *J. of Elec. Im.*, 6(1):94–107, Jan. 1997.

[8] A. Braun and S. R. Musse. Modeling individual behaviors in crowd simulation. In *Proc. of Comp. Anim. and Social Agents (*CASA'03*)*, New Jersey, USA, May 2003.

[9] D. Brogan and N. Johnson. Realistic human walking paths. In *Proc. of Comp. Anim. and Social Agents (*CASA'03*)*, pages 94–101, New Jersey, USA, May 2003.

[10] J. Chai, J. Xiao, and J. Hodgins. Vision-based control of 3d facial animation. In *Eurographics/ACM SIGGRAPH Symp. on Comp. Anim.*, pages 79–87, Grenoble, France, Aug. 2003.

[11] S. Chenney. Flow tiles. In *Eurographics/ACM SIGGRAPH Symp. on Comp. Anim. (SCA'04)*, pages 233–242, Grenoble, France, Aug. 2004.

[12] N. Courty and S. Musse. Simulation of Large Crowds Including Gaseous Phenomena. In *Proc. of IEEE Comp. Graph. Int. 2005*, pages 206–212, New York, USA, June 2005.

[13] D. Crisan. Particle filters, a theoretical perspective. In N. d. F. A. Doucet and N. Gordon, editors, *Sequential Monte-Carlo Methods in Practice*. Springer, 2001.

[14] J. Diener, L. Reveret, and E. Fiume. Hierarchical retargetting of 2d motion fields to the animation of 3d plant models. In *Eurographics/ACM SIGGRAPH Symp. on Comp. Anim. (SCA'06)*, Vienna, Austria, Aug. 2006.

[15] L. Favreau, L. Reveret, C. Depraz, and M.-P. Cani. Animal gaits from video. In *Eurographics/ACM SIGGRAPH Symp. on Comp. Anim. (SCA'04)*, Grenoble, France, Aug. 2004.

[16] D. Gibson, D. Oziem, C. Dalton, and N. Campbell. Capture and synthesis of insect motion. In *Eurographics/ACM SIGGRAPH Symp. on Comp. Anim. (SCA'05)*, pages 39–48, July 2005.

[17] D. Helbing, I. Farkas, and T. Vicsek. Simulating dynamical features of escape panic. *Nature*, 407(1):487–490, 2000.

[18] B. Horn and B. Schunck. Determining optical flow. *Art. Intell.*, 17:185–203, 1981.

[19] P. Huber. *Robust Statistics*. John Wiley & Sons, 1981.

[20] R. L. Hughes. The flow of human crowds. *Annual revue of Fluid. Mech.*, 20(10):169–182, 2003.

[21] L. Kovar, M. Gleicher, and F. Pighin. Motion graphs. *ACM Tran. on Graph., special issue, Proc. ACM SIGGRAPH 2002*, 21(3):473–482, 2002.

[22] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *In Proc. Seventh Int. Joint Conf. on Art. Intell.*, pages 674–679, Vancouver, Canada,, 1981.

[23] E. Mémin and P. Pérez. Dense estimation and object-based segmentation of the optical flow with robust techniques. *IEEE Trans. Im. Proc.*, 7(5):703–719, 1998.

[24] A. Mitiche and P. Bouthemy. Computation and analysis of image motion: a synopsis of current problems and methods. *Int. J. Comp. Vis.*, 19(1):29–55, 1996.

[25] S. R. Musse and D. Thalmann. Hierarchical model for real time simulation of virtual human crowds. In *IEEE Trans. on Vis and Comp. Graph.*, volume 7(2), pages 152–164. IEEE Comp. Society, 2001.

[26] T. Sakuma, T. Mukai, and S. Kuriyama. Psychological model for animating crowded pedestrians. *J. of Vis and Comp. Anim.*, 16(3-4):343–351, 2005.

[27] A. Schödl, R. Szeliski, D. Salesin, and I. Essa. Video textures. *ACM Tran. on Graph., special issue, Proc. ACM SIGGRAPH 2000*, pages 489–498, 2000.

[28] W. Shao and D. Terzopoulos. Animating autonomous pedestrians. In *Proc. SIGGRAPH/EG Symp. on Comp. Anim. (SCA'05)*, pages 19–28, Los Angeles, CA, July 2005.

[29] A. Somasundaram and R. Parent. Inserting synthetic characters into live-action scenes of multiple people. In *Proc. of Comp. Anim. and Social Agents (*CASA'03*)*, pages 137–142, New Jersey, USA, May 2003.

[30] M. Sung, M. Gleicher, and S. Chenney. Scalable behaviors for crowd simulation. *Comput. Graph. Forum*, 23(3):519–528, 2004.

[31] A. Treuille, S. Cooper, and Z. Popovic. Continuum crowds. *ACM Tran. on Graph., special issue, Proc. ACM SIGGRAPH 2006*, 25(3):1160–1168, 2006.

[32] T. Yang, S. Li, Q. Pan, and JingLi. Real-time multiple object tracking with occlusion handling in dynamic scenes. In *Proc. Conf. Comp. Vis. Patt. Rec.*, pages 406–413, San Diego, USA, June 2005.

[33] T. Zhao and R. Nevatia. Tracking multiple humans in crowded environment. In *Proc. Conf. Comp. Vis. Patt. Rec.*, pages 406–413, Washington, DC, USA, 2004.